Medical Image Data Mining for Diagnosis Support: Mammogram Case Studies

Barnes, Christopher F.
Georgia Institute of Technology, Savannah, GA, USA

Georgia Tech researchers have developed revolutionary image data mining systems initially designed to support non-medical applications such as remote sensing and automatic target recognition in defense systems. The core technologies of Georgia Tech's approach is highly innovative and theoretically sound, and have been under basic research for more than a decade. This basic research reached a milestone in the spring of 2002 with the development of a fully functional software prototype that runs on personal computers. Georgia Tech's image data mining tools can be easily tailored to support various applications. For example, image data mining can be used in the following areas:

- **Bioinformatics:** e.g., medical imaging and genome sequencing
- **Machine Vision:** e.g., manufacturing and maintenance inspection
- **Remote Sensing:** e.g., aerial and satellite image understanding
- **Security & Government Systems:** e.g., biometrics and intelligence

The key to Georgia Tech's image data mining is multimedia database technologies that surpass those in the best of today's commercial systems. Standard database tools are unable to effectively search non-traditional database information such as images. The image data mining tools built at Georgia Tech empower computers to fully explore the rich content of new multimedia databases. Georgia Tech researchers believe that data mining can find information beyond the capabilities of un-aided humans to analyze and understand medical images. Georgia Tech data mining increases the efficiency at which application experts are able to glean information from images; for example, provide additional information to radiologists to support diagnosis (increased accuracy rate) of suspected cancers by comparing a patient's mammogram with the most similar cases extracted from an archive of past mammograms. Knowledge base reasoning via image content data mining (KBR-ICD) is a new and revolutionary approach to detecting features and estimating the attributes of detected features in imagery. KBR-ICD's theoretical foundation is based on information theoretic, progressive, multidimensional reduced-degree-of-freedom data representations tightly integrated with standard relational database technologies. KBR-ICD enables systems to provide assisted and automated object, texture, clutter, and target recognition capabilities with orders of magnitude reduction in the complexity of other comparable adaptive-template approaches. Resulting knowledge base systems have the unique capability of providing both case-based content and context analysis. KBR-ICD empowers pixel-based image data mining for feature extraction and autonomous attribution of detected features in various types of imagery. KBR-ICD revolutionizes the concept of "text-hyperlinks contained in documents" to that of "pixel-hyperlinks contained in images." The linkage of these pixel-hyperlinks is directly derived from raw sensor signature data without requiring labor-intensive embedded metadata descriptions. Each pixel-hyperlink is a "clickable" object that an image analyst can "visit" or "data mine" to extract additional information from archived heterogeneous imagery and information.

Mammogram data mining research is currently being funded by the Georgia Cancer Coalition.